Student Feedback Survey Analysis

This notebook will walk you through the whole data cleaning, data analysis and insights discovery process of this workflow on a step-by-step basis. The steps are as follows:

1. Installation of Libraries.
2. Loading the data from file.
3. Cleaning and Transforming the data.
4. Exploratory Data Analysis (EDA).
5. Insights and Recommendations.
6. Export of cleaned data

1. Installation of Libraries

- The libraries needed for the analysis of the survey data are installed.

```
!pip install pandas numpy matplotlib seaborn wordcloud textblob
vadersentiment
```

```
Requirement already satisfied: pandas in
/usr/local/lib/python3.12/dist-packages (2.2.2)
Requirement already satisfied: numpy in
/usr/local/lib/python3.12/dist-packages (2.0.2)
Requirement already satisfied: matplotlib in
/usr/local/lib/python3.12/dist-packages (3.10.0)
Requirement already satisfied: seaborn in
/usr/local/lib/python3.12/dist-packages (0.13.2)
Requirement already satisfied: wordcloud in
/usr/local/lib/python3.12/dist-packages (1.9.4)
Requirement already satisfied: textblob in
/usr/local/lib/python3.12/dist-packages (0.19.0)
Collecting vadersentiment
  Downloading vaderSentiment-3.3.2-py2.py3-none-any.whl.metadata (572
bytes)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.12/dist-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in
/usr/local/lib/python3.12/dist-packages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in
/usr/local/lib/python3.12/dist-packages (from pandas) (2025.2)
Requirement already satisfied: contourpy>=1.0.1 in
/usr/local/lib/python3.12/dist-packages (from matplotlib) (1.3.3)
Requirement already satisfied: cycler>=0.10 in
/usr/local/lib/python3.12/dist-packages (from matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
/usr/local/lib/python3.12/dist-packages (from matplotlib) (4.60.1)
Requirement already satisfied: kiwisolver>=1.3.1 in
/usr/local/lib/python3.12/dist-packages (from matplotlib) (1.4.9)
```

```
Requirement already satisfied: packaging>=20.0 in
/usr/local/lib/python3.12/dist-packages (from matplotlib) (25.0)
Requirement already satisfied: pillow>=8 in
/usr/local/lib/python3.12/dist-packages (from matplotlib) (11.3.0)
Requirement already satisfied: pyparsing>=2.3.1 in
/usr/local/lib/python3.12/dist-packages (from matplotlib) (3.2.5)
Requirement already satisfied: nltk>=3.9 in
/usr/local/lib/python3.12/dist-packages (from textblob) (3.9.1)
Requirement already satisfied: requests in
/usr/local/lib/python3.12/dist-packages (from vadersentiment) (2.32.4)
Requirement already satisfied: click in
/usr/local/lib/python3.12/dist-packages (from nltk>=3.9->textblob)
(8.3.0)
Requirement already satisfied: joblib in
/usr/local/lib/python3.12/dist-packages (from nltk>=3.9->textblob)
(1.5.2)
Requirement already satisfied: regex>=2021.8.3 in
/usr/local/lib/python3.12/dist-packages (from nltk>=3.9->textblob)
(2024.11.6)
Requirement already satisfied: tqdm in /usr/local/lib/python3.12/dist-
packages (from nltk>=3.9->textblob) (4.67.1)
Requirement already satisfied: six>=1.5 in
/usr/local/lib/python3.12/dist-packages (from python-dateutil>=2.8.2-
>pandas) (1.17.0)
Requirement already satisfied: charset_normalizer<4,>=2 in
/usr/local/lib/python3.12/dist-packages (from requests-
>vadersentiment) (3.4.3)
Requirement already satisfied: idna<4,>=2.5 in
/usr/local/lib/python3.12/dist-packages (from requests-
>vadersentiment) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.12/dist-packages (from requests-
>vadersentiment) (2.5.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.12/dist-packages (from requests-
>vadersentiment) (2025.10.5)
Downloading vaderSentiment-3.3.2-py2.py3-none-any.whl (125 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 126.0/126.0 kB 7.7 MB/s eta
0:00:00
ent
Successfully installed vadersentiment-3.3.2
```

- The Natural Language Toolkit (NLTK) Library is downloaded and imported.

```
import nltk
nltk.download()

NLTK Downloader
---------------------------------------------------------------------
-----
```

```
    d) Download   l) List    u) Update   c) Config   h) Help   q) Quit
---------------------------------------------------------------------------
-----
Downloader> d

Download which package (l=list; x=cancel)?
  Identifier> l
Packages:
  [ ] abc................. Australian Broadcasting Commission 2006
  [ ] alpino.............. Alpino Dutch Treebank
  [ ] averaged_perceptron_tagger Averaged Perceptron Tagger
  [ ] averaged_perceptron_tagger_eng Averaged Perceptron Tagger (JSON)
  [ ] averaged_perceptron_tagger_ru Averaged Perceptron Tagger
(Russian)
  [ ] averaged_perceptron_tagger_rus Averaged Perceptron Tagger
(Russian)
  [ ] basque_grammars..... Grammars for Basque
  [ ] bcp47............... BCP-47 Language Tags
  [ ] biocreative_ppi..... BioCreAtIvE (Critical Assessment of
Information
                           Extraction Systems in Biology)
  [ ] bllip_wsj_no_aux.... BLLIP Parser: WSJ Model
  [ ] book_grammars....... Grammars from NLTK Book
  [ ] brown............... Brown Corpus
  [ ] brown_tei........... Brown Corpus (TEI XML Version)
  [ ] cess_cat............ CESS-CAT Treebank
  [ ] cess_esp............ CESS-ESP Treebank
  [ ] chat80.............. Chat-80 Data Files
  [ ] city_database....... City Database
  [ ] cmudict............. The Carnegie Mellon Pronouncing Dictionary
(0.6)
  [ ] comparative_sentences Comparative Sentence Dataset
Hit Enter to continue:
  [ ] comtrans............ ComTrans Corpus Sample
  [ ] conll2000........... CONLL 2000 Chunking Corpus
  [ ] conll2002........... CONLL 2002 Named Entity Recognition Corpus
  [ ] conll2007........... Dependency Treebanks from CoNLL 2007
(Catalan
                           and Basque Subset)
  [ ] crubadan............ Crubadan Corpus
  [ ] dependency_treebank. Dependency Parsed Treebank
  [ ] dolch............... Dolch Word List
  [ ] english_wordnet..... Open English Wordnet
  [ ] europarl_raw........ Sample European Parliament Proceedings
Parallel
                           Corpus
  [ ] extended_omw........ Extended Open Multilingual WordNet
  [ ] floresta............ Portuguese Treebank
  [ ] framenet_v15........ FrameNet 1.5
```

```
  [ ] framenet_v17........ FrameNet 1.7
  [ ] gazetteers.......... Gazeteer Lists
  [ ] genesis............. Genesis Corpus
  [ ] gutenberg........... Project Gutenberg Selections
  [ ] ieer................ NIST IE-ER DATA SAMPLE
  [ ] inaugural........... C-Span Inaugural Address Corpus
  [ ] indian.............. Indian Language POS-Tagged Corpus
Hit Enter to continue:
  [ ] jeita............... JEITA Public Morphologically Tagged Corpus
(in
                           ChaSen format)
  [ ] kimmo............... PC-KIMMO Data Files
  [ ] knbc................ KNB Corpus (Annotated blog corpus)
  [ ] large_grammars...... Large context-free and feature-based
grammars
                           for parser comparison
  [ ] lin_thesaurus....... Lin's Dependency Thesaurus
  [ ] mac_morpho.......... MAC-MORPHO: Brazilian Portuguese news text
with
                           part-of-speech tags
  [ ] machado............. Machado de Assis -- Obra Completa
  [ ] masc_tagged......... MASC Tagged Corpus
  [ ] maxent_ne_chunker... ACE Named Entity Chunker (Maximum entropy)
  [ ] maxent_ne_chunker_tab ACE Named Entity Chunker (Maximum entropy)
  [ ] maxent_treebank_pos_tagger Treebank Part of Speech Tagger
(Maximum entropy)
  [ ] maxent_treebank_pos_tagger_tab Treebank Part of Speech Tagger
(Maximum entropy)
  [ ] mock_corpus......... Mock Corpus
  [ ] moses_sample........ Moses Sample Models
  [ ] movie_reviews....... Sentiment Polarity Dataset Version 2.0
  [ ] mte_teip5........... MULTEXT-East 1984 annotated corpus 4.0
  [ ] mwa_ppdb............ The monolingual word aligner (Sultan et al.
                           2015) subset of the Paraphrase Database.
Hit Enter to continue:
  [ ] names............... Names Corpus, Version 1.3 (1994-03-29)
  [ ] nombank.1.0......... NomBank Corpus 1.0
  [ ] nonbreaking_prefixes Non-Breaking Prefixes (Moses Decoder)
  [ ] nps_chat............ NPS Chat
  [ ] omw-1.4............. Open Multilingual Wordnet
  [ ] omw................. Open Multilingual Wordnet
  [ ] opinion_lexicon..... Opinion Lexicon
  [ ] panlex_swadesh...... PanLex Swadesh Corpora
  [ ] paradigms........... Paradigm Corpus
  [ ] pe08................ Cross-Framework and Cross-Domain Parser
                           Evaluation Shared Task
  [ ] perluniprops........ perluniprops: Index of Unicode Version
7.0.0
                           character properties in Perl
```

```
  [ ] pil................. The Patient Information Leaflet (PIL)
Corpus
  [ ] pl196x.............. Polish language of the XX century sixties
  [ ] porter_test......... Porter Stemmer Test Files
  [ ] ppattach............ Prepositional Phrase Attachment Corpus
  [ ] problem_reports..... Problem Report Corpus
  [ ] product_reviews_1... Product Reviews (5 Products)
  [ ] product_reviews_2... Product Reviews (9 Products)
  [ ] propbank............ Proposition Bank Corpus 1.0
Hit Enter to continue:
  [ ] pros_cons........... Pros and Cons
  [ ] ptb................. Penn Treebank
  [ ] punkt............... Punkt Tokenizer Models
  [ ] punkt_tab........... Punkt Tokenizer Models
  [ ] qc.................. Experimental Data for Question
Classification
  [ ] reuters............. The Reuters-21578 benchmark corpus, ApteMod
                           version
  [ ] rslp................ RSLP Stemmer (Removedor de Sufixos da
Lingua
                           Portuguesa)
  [ ] rte................. PASCAL RTE Challenges 1, 2, and 3
  [ ] sample_grammars..... Sample Grammars
  [ ] semcor.............. SemCor 3.0
  [ ] senseval............ SENSEVAL 2 Corpus: Sense Tagged Text
  [ ] sentence_polarity... Sentence Polarity Dataset v1.0
  [ ] sentiwordnet........ SentiWordNet
  [ ] shakespeare......... Shakespeare XML Corpus Sample
  [ ] sinica_treebank..... Sinica Treebank Corpus Sample
  [ ] smultron............ SMULTRON Corpus Sample
  [ ] snowball_data....... Snowball Data
  [ ] spanish_grammars.... Grammars for Spanish
  [ ] state_union......... C-Span State of the Union Address Corpus
Hit Enter to continue:
  [ ] stopwords........... Stopwords Corpus
  [ ] subjectivity........ Subjectivity Dataset v1.0
  [ ] swadesh............. Swadesh Wordlists
  [ ] switchboard......... Switchboard Corpus Sample
  [ ] tagsets............. Help on Tagsets
  [ ] tagsets_json........ Help on Tagsets (JSON)
  [ ] timit............... TIMIT Corpus Sample
  [ ] toolbox............. Toolbox Sample Files
  [ ] treebank............ Penn Treebank Sample
  [ ] twitter_samples..... Twitter Samples
  [ ] udhr2............... Universal Declaration of Human Rights
Corpus
                           (Unicode Version)
  [ ] udhr................ Universal Declaration of Human Rights
Corpus
```

```
  [ ] unicode_samples..... Unicode Samples
  [ ] universal_tagset.... Mappings to the Universal Part-of-Speech
Tagset
  [ ] universal_treebanks_v20 Universal Treebanks Version 2.0
  [ ] vader_lexicon....... VADER Sentiment Lexicon
  [ ] verbnet3............ VerbNet Lexicon, Version 3.3
  [ ] verbnet............. VerbNet Lexicon, Version 2.1
  [ ] webtext............. Web Text Corpus
  [ ] wmt15_eval.......... Evaluation data from WMT15
Hit Enter to continue:
  [ ] word2vec_sample..... Word2Vec Sample
  [ ] wordnet2021......... Open English Wordnet 2021
  [ ] wordnet2022......... Open English Wordnet 2022
  [ ] wordnet31........... Wordnet 3.1
  [ ] wordnet............. WordNet
  [ ] wordnet_ic.......... WordNet-InfoContent
  [ ] words............... Word Lists
  [ ] ycoe................ York-Toronto-Helsinki Parsed Corpus of Old
                           English Prose

Collections:
  [ ] all-corpora......... All the corpora
  [ ] all-nltk............ All packages available on nltk_data gh-
pages
                           branch
  [ ] all................. All packages
  [ ] book................ Everything used in the NLTK Book
  [ ] popular............. Popular packages
  [ ] tests............... Packages for running tests
  [ ] third-party......... Third-party data packages

([*] marks installed packages)

Download which package (l=list; x=cancel)?
  Identifier> all

    Downloading collection 'all'
       |
       | Downloading package abc to /root/nltk_data...
       |   Unzipping corpora/abc.zip.
       | Downloading package alpino to /root/nltk_data...
       |   Unzipping corpora/alpino.zip.
       | Downloading package averaged_perceptron_tagger to
       |     /root/nltk_data...
       |   Unzipping taggers/averaged_perceptron_tagger.zip.
       | Downloading package averaged_perceptron_tagger_eng to
       |     /root/nltk_data...
       |   Unzipping taggers/averaged_perceptron_tagger_eng.zip.
       | Downloading package averaged_perceptron_tagger_ru to
       |     /root/nltk_data...
```

```
[nltk_data]    |   Unzipping taggers/averaged_perceptron_tagger_ru.zip.
[nltk_data]    | Downloading package averaged_perceptron_tagger_rus to
[nltk_data]    |     /root/nltk_data...
[nltk_data]    |   Unzipping taggers/averaged_perceptron_tagger_rus.zip.
[nltk_data]    | Downloading package basque_grammars to /root/nltk_data...
[nltk_data]    |   Unzipping grammars/basque_grammars.zip.
[nltk_data]    | Downloading package bcp47 to /root/nltk_data...
[nltk_data]    | Downloading package biocreative_ppi to /root/nltk_data...
[nltk_data]    |   Unzipping corpora/biocreative_ppi.zip.
[nltk_data]    | Downloading package bllip_wsj_no_aux to /root/nltk_data...
[nltk_data]    |   Unzipping models/bllip_wsj_no_aux.zip.
[nltk_data]    | Downloading package book_grammars to /root/nltk_data...
[nltk_data]    |   Unzipping grammars/book_grammars.zip.
[nltk_data]    | Downloading package brown to /root/nltk_data...
[nltk_data]    |   Unzipping corpora/brown.zip.
[nltk_data]    | Downloading package brown_tei to /root/nltk_data...
[nltk_data]    |   Unzipping corpora/brown_tei.zip.
[nltk_data]    | Downloading package cess_cat to /root/nltk_data...
[nltk_data]    |   Unzipping corpora/cess_cat.zip.
[nltk_data]    | Downloading package cess_esp to /root/nltk_data...
[nltk_data]    |   Unzipping corpora/cess_esp.zip.
[nltk_data]    | Downloading package chat80 to /root/nltk_data...
[nltk_data]    |   Unzipping corpora/chat80.zip.
[nltk_data]    | Downloading package city_database to /root/nltk_data...
[nltk_data]    |   Unzipping corpora/city_database.zip.
[nltk_data]    | Downloading package cmudict to /root/nltk_data...
[nltk_data]    |   Unzipping corpora/cmudict.zip.
[nltk_data]    | Downloading package comparative_sentences to
[nltk_data]    |     /root/nltk_data...
[nltk_data]    |   Unzipping corpora/comparative_sentences.zip.
[nltk_data]    | Downloading package comtrans to /root/nltk_data...
[nltk_data]    | Downloading package conll2000 to /root/nltk_data...
[nltk_data]    |   Unzipping corpora/conll2000.zip.
[nltk_data]    | Downloading package conll2002 to /root/nltk_data...
[nltk_data]    |   Unzipping corpora/conll2002.zip.
[nltk_data]    | Downloading package conll2007 to /root/nltk_data...
[nltk_data]    | Downloading package crubadan to /root/nltk_data...
[nltk_data]    |   Unzipping corpora/crubadan.zip.
[nltk_data]    | Downloading package dependency_treebank to /root/nltk_data...
[nltk_data]    |   Unzipping corpora/dependency_treebank.zip.
[nltk_data]    | Downloading package dolch to /root/nltk_data...
[nltk_data]    |   Unzipping corpora/dolch.zip.
[nltk_data]    | Downloading package english_wordnet to /root/nltk_data...
[nltk_data]    |   Unzipping corpora/english_wordnet.zip.
[nltk_data]    | Downloading package europarl_raw to /root/nltk_data...
[nltk_data]    |   Unzipping corpora/europarl_raw.zip.
[nltk_data]    | Downloading package extended_omw to /root/nltk_data...
[nltk_data]    | Downloading package floresta to /root/nltk_data...
[nltk_data]    |   Unzipping corpora/floresta.zip.
```

```
       | Downloading package framenet_v15 to /root/nltk_data...
       |   Unzipping corpora/framenet_v15.zip.
       | Downloading package framenet_v17 to /root/nltk_data...
       |   Unzipping corpora/framenet_v17.zip.
       | Downloading package gazetteers to /root/nltk_data...
       |   Unzipping corpora/gazetteers.zip.
       | Downloading package genesis to /root/nltk_data...
       |   Unzipping corpora/genesis.zip.
       | Downloading package gutenberg to /root/nltk_data...
       |   Unzipping corpora/gutenberg.zip.
       | Downloading package ieer to /root/nltk_data...
       |   Unzipping corpora/ieer.zip.
       | Downloading package inaugural to /root/nltk_data...
       |   Unzipping corpora/inaugural.zip.
       | Downloading package indian to /root/nltk_data...
       |   Unzipping corpora/indian.zip.
       | Downloading package jeita to /root/nltk_data...
       | Downloading package kimmo to /root/nltk_data...
       |   Unzipping corpora/kimmo.zip.
       | Downloading package knbc to /root/nltk_data...
       | Downloading package large_grammars to /root/nltk_data...
       |   Unzipping grammars/large_grammars.zip.
       | Downloading package lin_thesaurus to /root/nltk_data...
       |   Unzipping corpora/lin_thesaurus.zip.
       | Downloading package mac_morpho to /root/nltk_data...
       |   Unzipping corpora/mac_morpho.zip.
       | Downloading package machado to /root/nltk_data...
       | Downloading package masc_tagged to /root/nltk_data...
       | Downloading package maxent_ne_chunker to /root/nltk_data...
       |   Unzipping chunkers/maxent_ne_chunker.zip.
       | Downloading package maxent_ne_chunker_tab to
       |     /root/nltk_data...
       |   Unzipping chunkers/maxent_ne_chunker_tab.zip.
       | Downloading package maxent_treebank_pos_tagger to
       |     /root/nltk_data...
       |   Unzipping taggers/maxent_treebank_pos_tagger.zip.
       | Downloading package maxent_treebank_pos_tagger_tab to
       |     /root/nltk_data...
       |   Unzipping taggers/maxent_treebank_pos_tagger_tab.zip.
       | Downloading package mock_corpus to /root/nltk_data...
       |   Unzipping corpora/mock_corpus.zip.
       | Downloading package moses_sample to /root/nltk_data...
       |   Unzipping models/moses_sample.zip.
       | Downloading package movie_reviews to /root/nltk_data...
       |   Unzipping corpora/movie_reviews.zip.
       | Downloading package mte_teip5 to /root/nltk_data...
       |   Unzipping corpora/mte_teip5.zip.
       | Downloading package mwa_ppdb to /root/nltk_data...
       |   Unzipping misc/mwa_ppdb.zip.
```

```
[nltk_data] Downloading package names to /root/nltk_data...
[nltk_data]   Unzipping corpora/names.zip.
[nltk_data] Downloading package nombank.1.0 to /root/nltk_data...
[nltk_data] Downloading package nonbreaking_prefixes to
[nltk_data]     /root/nltk_data...
[nltk_data]   Unzipping corpora/nonbreaking_prefixes.zip.
[nltk_data] Downloading package nps_chat to /root/nltk_data...
[nltk_data]   Unzipping corpora/nps_chat.zip.
[nltk_data] Downloading package omw to /root/nltk_data...
[nltk_data] Downloading package omw-1.4 to /root/nltk_data...
[nltk_data] Downloading package opinion_lexicon to /root/nltk_data...
[nltk_data]   Unzipping corpora/opinion_lexicon.zip.
[nltk_data] Downloading package panlex_swadesh to /root/nltk_data...
[nltk_data] Downloading package paradigms to /root/nltk_data...
[nltk_data]   Unzipping corpora/paradigms.zip.
[nltk_data] Downloading package pe08 to /root/nltk_data...
[nltk_data]   Unzipping corpora/pe08.zip.
[nltk_data] Downloading package perluniprops to /root/nltk_data...
[nltk_data]   Unzipping misc/perluniprops.zip.
[nltk_data] Downloading package pil to /root/nltk_data...
[nltk_data]   Unzipping corpora/pil.zip.
[nltk_data] Downloading package pl196x to /root/nltk_data...
[nltk_data]   Unzipping corpora/pl196x.zip.
[nltk_data] Downloading package porter_test to /root/nltk_data...
[nltk_data]   Unzipping stemmers/porter_test.zip.
[nltk_data] Downloading package ppattach to /root/nltk_data...
[nltk_data]   Unzipping corpora/ppattach.zip.
[nltk_data] Downloading package problem_reports to /root/nltk_data...
[nltk_data]   Unzipping corpora/problem_reports.zip.
[nltk_data] Downloading package product_reviews_1 to /root/nltk_data...
[nltk_data]   Unzipping corpora/product_reviews_1.zip.
[nltk_data] Downloading package product_reviews_2 to /root/nltk_data...
[nltk_data]   Unzipping corpora/product_reviews_2.zip.
[nltk_data] Downloading package propbank to /root/nltk_data...
[nltk_data] Downloading package pros_cons to /root/nltk_data...
[nltk_data]   Unzipping corpora/pros_cons.zip.
[nltk_data] Downloading package ptb to /root/nltk_data...
[nltk_data]   Unzipping corpora/ptb.zip.
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
[nltk_data] Downloading package punkt_tab to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt_tab.zip.
[nltk_data] Downloading package qc to /root/nltk_data...
[nltk_data]   Unzipping corpora/qc.zip.
[nltk_data] Downloading package reuters to /root/nltk_data...
[nltk_data] Downloading package rslp to /root/nltk_data...
[nltk_data]   Unzipping stemmers/rslp.zip.
[nltk_data] Downloading package rte to /root/nltk_data...
[nltk_data]   Unzipping corpora/rte.zip.
```

```
       | Downloading package sample_grammars to /root/nltk_data...
       |   Unzipping grammars/sample_grammars.zip.
       | Downloading package semcor to /root/nltk_data...
       | Downloading package senseval to /root/nltk_data...
       |   Unzipping corpora/senseval.zip.
       | Downloading package sentence_polarity to /root/nltk_data...
       |   Unzipping corpora/sentence_polarity.zip.
       | Downloading package sentiwordnet to /root/nltk_data...
       |   Unzipping corpora/sentiwordnet.zip.
       | Downloading package shakespeare to /root/nltk_data...
       |   Unzipping corpora/shakespeare.zip.
       | Downloading package sinica_treebank to /root/nltk_data...
       |   Unzipping corpora/sinica_treebank.zip.
       | Downloading package smultron to /root/nltk_data...
       |   Unzipping corpora/smultron.zip.
       | Downloading package snowball_data to /root/nltk_data...
       | Downloading package spanish_grammars to /root/nltk_data...
       |   Unzipping grammars/spanish_grammars.zip.
       | Downloading package state_union to /root/nltk_data...
       |   Unzipping corpora/state_union.zip.
       | Downloading package stopwords to /root/nltk_data...
       |   Unzipping corpora/stopwords.zip.
       | Downloading package subjectivity to /root/nltk_data...
       |   Unzipping corpora/subjectivity.zip.
       | Downloading package swadesh to /root/nltk_data...
       |   Unzipping corpora/swadesh.zip.
       | Downloading package switchboard to /root/nltk_data...
       |   Unzipping corpora/switchboard.zip.
       | Downloading package tagsets to /root/nltk_data...
       |   Unzipping help/tagsets.zip.
       | Downloading package tagsets_json to /root/nltk_data...
       |   Unzipping help/tagsets_json.zip.
       | Downloading package timit to /root/nltk_data...
       |   Unzipping corpora/timit.zip.
       | Downloading package toolbox to /root/nltk_data...
       |   Unzipping corpora/toolbox.zip.
       | Downloading package treebank to /root/nltk_data...
       |   Unzipping corpora/treebank.zip.
       | Downloading package twitter_samples to /root/nltk_data...
       |   Unzipping corpora/twitter_samples.zip.
       | Downloading package udhr to /root/nltk_data...
       |   Unzipping corpora/udhr.zip.
       | Downloading package udhr2 to /root/nltk_data...
       |   Unzipping corpora/udhr2.zip.
       | Downloading package unicode_samples to /root/nltk_data...
       |   Unzipping corpora/unicode_samples.zip.
       | Downloading package universal_tagset to /root/nltk_data...
       |   Unzipping taggers/universal_tagset.zip.
       | Downloading package universal_treebanks_v20 to
```

```
       |     /root/nltk_data...
       | Downloading package vader_lexicon to /root/nltk_data...
       | Downloading package verbnet to /root/nltk_data...
       |   Unzipping corpora/verbnet.zip.
       | Downloading package verbnet3 to /root/nltk_data...
       |   Unzipping corpora/verbnet3.zip.
       | Downloading package webtext to /root/nltk_data...
       |   Unzipping corpora/webtext.zip.
       | Downloading package wmt15_eval to /root/nltk_data...
       |   Unzipping models/wmt15_eval.zip.
       | Downloading package word2vec_sample to /root/nltk_data...
       |   Unzipping models/word2vec_sample.zip.
       | Downloading package wordnet to /root/nltk_data...
       | Downloading package wordnet2021 to /root/nltk_data...
       | Downloading package wordnet2022 to /root/nltk_data...
       |   Unzipping corpora/wordnet2022.zip.
       | Downloading package wordnet31 to /root/nltk_data...
       | Downloading package wordnet_ic to /root/nltk_data...
       |   Unzipping corpora/wordnet_ic.zip.
       | Downloading package words to /root/nltk_data...
       |   Unzipping corpora/words.zip.
       | Downloading package ycoe to /root/nltk_data...
       |   Unzipping corpora/ycoe.zip.
       |
     Done downloading collection all


---------------------------------------------------------------------------
-----
    d) Download   l) List    u) Update   c) Config   h) Help   q) Quit
---------------------------------------------------------------------------
-----
Downloader> q
```

2. Loading the data from file

- Since the file containing the data to be analyzed is located in Google Drive, we first mount the drive and specify the path to the file in order to locate it.

```
from google.colab import drive
drive.mount('/content/drive')
import pandas as pd
file_path = '/content/drive/MyDrive/student_feedback.csv'
df = pd.read_csv(file_path)
df.head()
```

```
Drive already mounted at /content/drive; to attempt to forcibly
remount, call drive.mount("/content/drive", force_remount=True).
```

```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 1001,\n  \"fields\":
[\n    {\n      \"column\": \"Unnamed: 0\",\n      \"properties\": {\n
```

```
\"dtype\": \"number\",\n        \"std\": 289,\n        \"min\": 0,\n
\"max\": 1000,\n        \"num_unique_values\": 1001,\n
\"samples\": [\n          521,\n          941,\n          741\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Student ID\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
289,\n        \"min\": 0,\n        \"max\": 1000,\n
\"num_unique_values\": 1001,\n        \"samples\": [\n          765,\n
677,\n          537\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Well versed with the subject\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 1,\n        \"min\": 5,\n
\"max\": 10,\n        \"num_unique_values\": 6,\n        \"samples\":
[\n          5,\n          6,\n          10\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Explains concepts in an
understandable way\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 2,\n        \"min\": 2,\n
\"max\": 10,\n        \"num_unique_values\": 9,\n        \"samples\":
[\n          9,\n          5,\n          3\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Use of presentations\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
1,\n        \"min\": 4,\n        \"max\": 8,\n
\"num_unique_values\": 5,\n        \"samples\": [\n          8,\n
4,\n          6\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Degree of difficulty of assignments\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 2,\n        \"min\": 1,\n
\"max\": 10,\n        \"num_unique_values\": 10,\n        \"samples\":
[\n          10,\n          5,\n          2\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Solves doubts willingly\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
2,\n        \"min\": 1,\n        \"max\": 10,\n
\"num_unique_values\": 10,\n        \"samples\": [\n          1,\n
2,\n          3\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Structuring of the course\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 2,\n        \"min\": 1,\n
\"max\": 10,\n        \"num_unique_values\": 10,\n        \"samples\":
[\n          10,\n          1,\n          3\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Provides support for students
going above and beyond\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 2,\n        \"min\": 1,\n
\"max\": 10,\n        \"num_unique_values\": 10,\n        \"samples\":
[\n          10,\n          2,\n          8\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
```

```
n    },\n    {\n      \"column\": \"Course recommendation based on
relevance\",\n      \"properties\": {\n        \"dtype\": \"number\",\
n        \"std\": 2,\n        \"min\": 1,\n        \"max\": 10,\n
\"num_unique_values\": 10,\n        \"samples\": [\n          7,\n
9,\n          10\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    }\n  ]\
n}","type":"dataframe","variable_name":"df"}
```


3. Cleaning and Transforming the data

- First we check to see if any null values are present in the data

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1001 entries, 0 to 1000
Data columns (total 10 columns):
 #   Column                                                Non-Null
Count  Dtype
---  ------
--------------  -----
 0   Unnamed: 0                                            1001 non-
null   int64
 1   Student ID                                            1001 non-
null   int64
 2   Well versed with the subject                          1001 non-
null   int64
 3   Explains concepts in an understandable way            1001 non-
null   int64
 4   Use of presentations                                  1001 non-
null   int64
 5   Degree of difficulty of assignments                   1001 non-
null   int64
 6   Solves doubts willingly                               1001 non-
null   int64
 7   Structuring of the course                             1001 non-
null   int64
 8   Provides support for students going above and beyond  1001 non-
null   int64
 9   Course recommendation based on relevance              1001 non-
null   int64
dtypes: int64(10)
memory usage: 78.3 KB
```

No null values in data so we proceed to import the 'numpy' library and perform the following operations:

- Normalize column names
- Rename column names to common schema
- Coerce numeric parameters

- Sort missing numeric parameters
- Cleaning duplicates

```
import numpy as np

df = df.copy(deep=True)

# Normalize column names
df.columns = df.columns.str.strip().str.lower().str.replace('[^a-z0-
9]+','_', regex=True)

# Rename to common schema
rename_map = {
    'unnamed_0': 'row',
    'student_id': 'student_id',
    'well_versed_with_the_subject': 'well_versed_with_the_subject',
    'explains_concepts_in_an_understandable_way':
'explains_concepts_in_an_understandable_way',
    'use_of_presentations': 'use_of_presentations',
    'degree_of_difficulty_of_assignments':
'degree_of_difficulty_of_assignments',
    'solves_doubts_willingly': 'solves_doubts_willingly',
    'structuring_of_the_course': 'structuring_of_the_course',
    'provides_support_for_students_going_above_and_beyond':
'provides_support_for_students_going_above_and_beyond',
    'course_recommendation_based_on_relevance':
'course_recommendation_based_on_relevance',
}
df = df.rename(columns={k:v for k,v in rename_map.items() if k in
df.columns})

# Coerce numeric values
for col in ['well_versed_with_the_subject',
'explains_concepts_in_an_understandable_way', 'use_of_presentations',
'degree_of_difficulty_of_assignments', 'solves_doubts_willingly',
'structuring_of_the_course',
'provides_support_for_students_going_above_and_beyond',
'course_recommendation_based_on_relevance']:
  if col in df.columns:
    df[col] = pd.to_numeric(df[col], errors='coerce')

# Sort missing numerical parameters
if ('well_versed_with_the_subject',
'explains_concepts_in_an_understandable_way', 'use_of_presentations',
'degree_of_difficulty_of_assignments', 'solves_doubts_willingly',
'structuring_of_the_course',
'provides_support_for_students_going_above_and_beyond',
'course_recommendation_based_on_relevance') in df.columns:
    df = df[df['well_versed_with_the_subject',
'explains_concepts_in_an_understandable_way', 'use_of_presentations',
```

```
'degree_of_difficulty_of_assignments', 'solves_doubts_willingly',
'structuring_of_the_course',
'provides_support_for_students_going_above_and_beyond',
'course_recommendation_based_on_relevance'].between(1,10)]

#Cleaning duplicates
df = df.drop_duplicates()

print(df.shape)
df.head()
```

```
(1001, 10)
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 1001,\n  \"fields\":
[\n    {\n      \"column\": \"row\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 289,\n        \"min\": 0,\n
\"max\": 1000,\n        \"num_unique_values\": 1001,\n
\"samples\": [\n          521,\n          941,\n          741\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"student_id\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
289,\n        \"min\": 0,\n        \"max\": 1000,\n
\"num_unique_values\": 1001,\n        \"samples\": [\n          765,\n
677,\n          537\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"well_versed_with_the_subject\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 1,\n        \"min\": 5,\n
\"max\": 10,\n        \"num_unique_values\": 6,\n        \"samples\":
[\n          5,\n          6,\n          10\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\":
\"explains_concepts_in_an_understandable_way\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\": 2,\n
\"min\": 2,\n        \"max\": 10,\n        \"num_unique_values\": 9,\n
\"samples\": [\n          9,\n          5,\n          3\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"use_of_presentations\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
1,\n        \"min\": 4,\n        \"max\": 8,\n
\"num_unique_values\": 5,\n        \"samples\": [\n          8,\n
4,\n          6\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"degree_of_difficulty_of_assignments\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 2,\n        \"min\": 1,\n
\"max\": 10,\n        \"num_unique_values\": 10,\n        \"samples\":
[\n          10,\n          5,\n          2\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"solves_doubts_willingly\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
2,\n        \"min\": 1,\n        \"max\": 10,\n
```

```
\"num_unique_values\": 10,\n        \"samples\": [\n          1,\n
2,\n          3\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"structuring_of_the_course\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 2,\n        \"min\": 1,\n
\"max\": 10,\n        \"num_unique_values\": 10,\n        \"samples\":
[\n          10,\n          1,\n          3\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\":
\"provides_support_for_students_going_above_and_beyond\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
2,\n        \"min\": 1,\n        \"max\": 10,\n
\"num_unique_values\": 10,\n        \"samples\": [\n          10,\n
2,\n          8\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"course_recommendation_based_on_relevance\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\": 2,\n
\"min\": 1,\n        \"max\": 10,\n        \"num_unique_values\": 10,\
n        \"samples\": [\n          7,\n          9,\n          10\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    }\n  ]\n}","type":"dataframe","variable_name":"df"}
```


4. Exploratory Data Analysis (EDA)

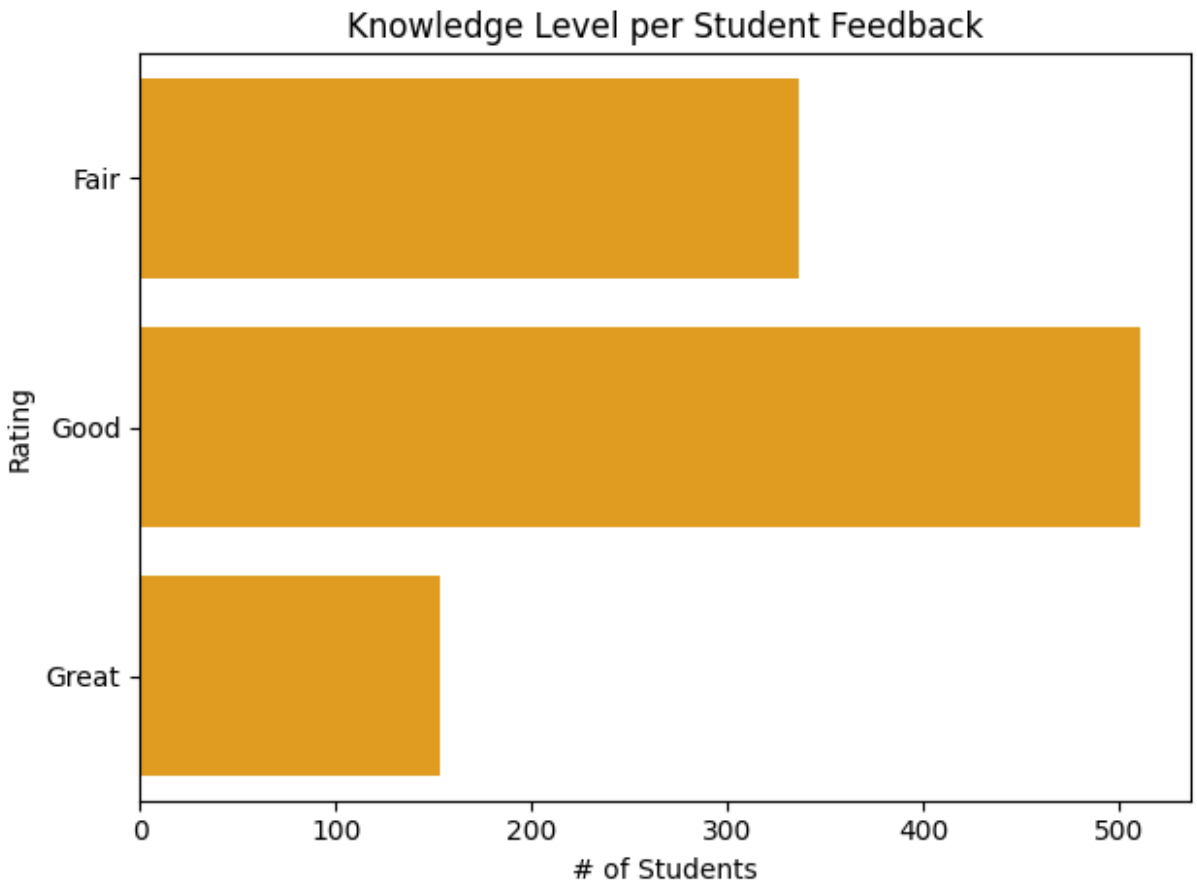
In this section, we first of all created bin labels to group the numerical ratings into 3 categories: 'Fair', 'Good' and 'Great' for easy reference. After this was done, we then proceeded to uncover the following insights:

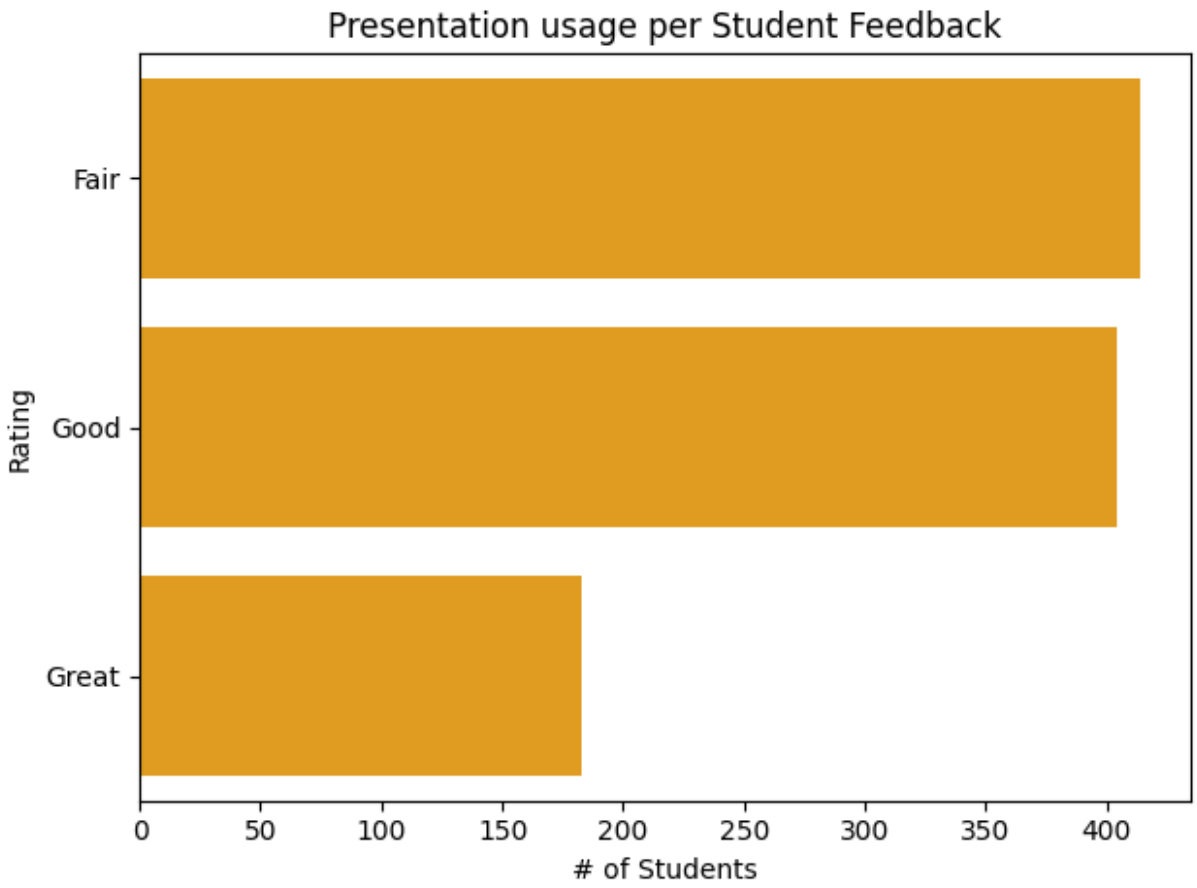
- Knowledge level rating per student feedback.
- Presentation rating per student feedback.
- Concepts explanation per student feedback.
- Assignment difficulty per student feedback.
- Doubt resolution per student feedback.
- Course structure per student feedback.
- Student support per student feedback.
- Course recommendation per student feedback.

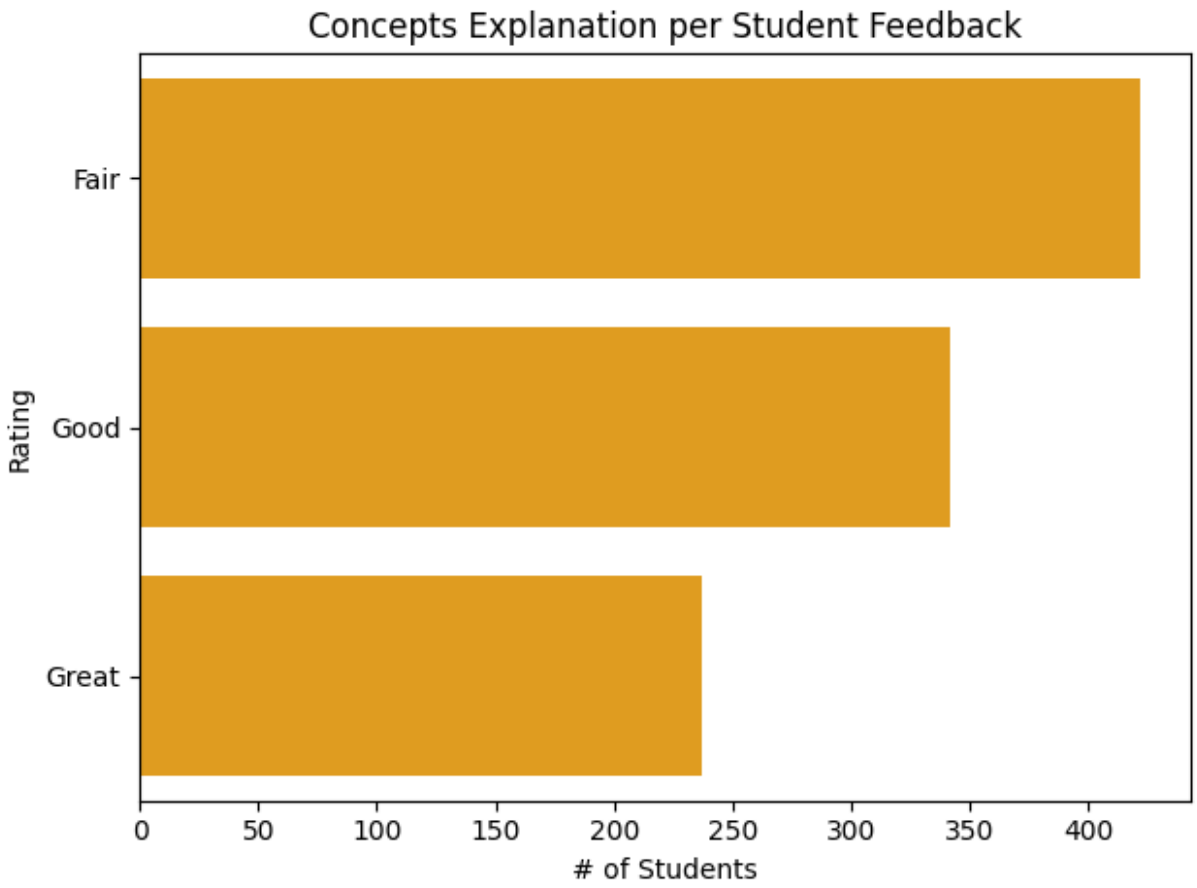
```
import matplotlib.pyplot as plt
import seaborn as sns

#Creation of Bin Labels for rating categories
bin_labels = ['Fair', 'Good', 'Great']

#Knowledge level rating per student feedback
well_versed_with_the_subject_group =
pd.qcut(df.well_versed_with_the_subject,labels=bin_labels,q=3)
ax = sns.countplot(well_versed_with_the_subject_group, color='orange')
ax.set_title('Knowledge Level per Student Feedback')
ax.set_xlabel('# of Students')
```

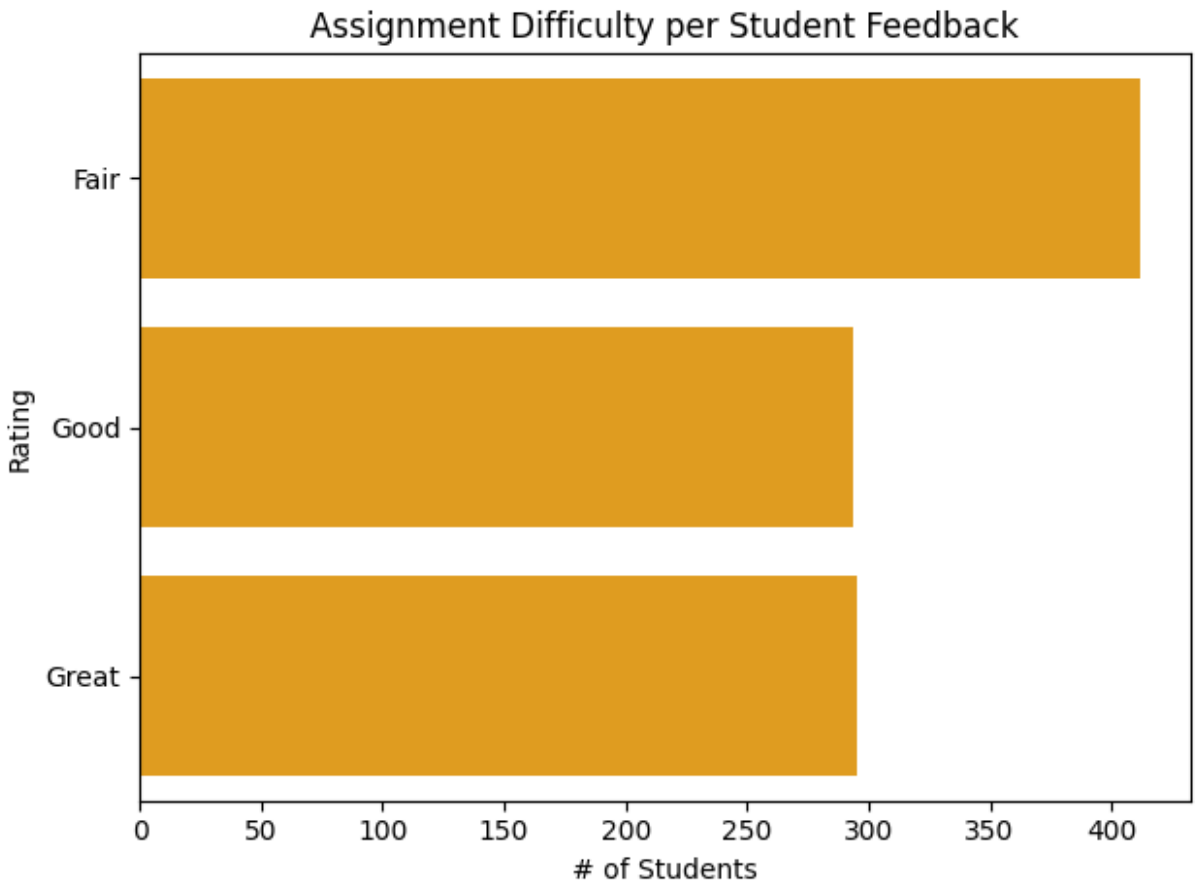
```
ax.set_ylabel('Rating')
plt.tight_layout()
plt.show()

#Presentation rating per student feedback
use_of_presentations_group =
pd.qcut(df.use_of_presentations,labels=bin_labels,q=3)
ax = sns.countplot(use_of_presentations_group, color='orange')
ax.set_title('Presentation usage per Student Feedback')
ax.set_xlabel('# of Students')
ax.set_ylabel('Rating')
plt.tight_layout()
plt.show()

#Rating for explanation of concepts per student feedback
explains_concepts_in_an_understandable_way_group =
pd.qcut(df.explains_concepts_in_an_understandable_way,labels=bin_label
s,q=3)
ax = sns.countplot(explains_concepts_in_an_understandable_way_group,
color='orange')
ax.set_title('Concepts Explanation per Student Feedback')
ax.set_xlabel('# of Students')
ax.set_ylabel('Rating')
plt.tight_layout()
plt.show()

#Assignment difficulty rating per student feedback
degree_of_difficulty_of_assignments_group =
pd.qcut(df.degree_of_difficulty_of_assignments,labels=bin_labels,q=3)
ax = sns.countplot(degree_of_difficulty_of_assignments_group,
color='orange')
ax.set_title('Assignment Difficulty per Student Feedback')
ax.set_xlabel('# of Students')
ax.set_ylabel('Rating')
plt.tight_layout()
plt.show()

#Doubt resolution rating per student feedback
solves_doubts_willingly_group =
pd.qcut(df.solves_doubts_willingly,labels=bin_labels,q=3)
ax = sns.countplot(solves_doubts_willingly_group, color='orange')
ax.set_title('Doubt Resolution per Student Feedback')
ax.set_xlabel('# of Students')
ax.set_ylabel('Rating')
plt.tight_layout()
plt.show()

#Course Structure rating per student feedback
structuring_of_the_course_group =
pd.qcut(df.structuring_of_the_course,labels=bin_labels,q=3)
```

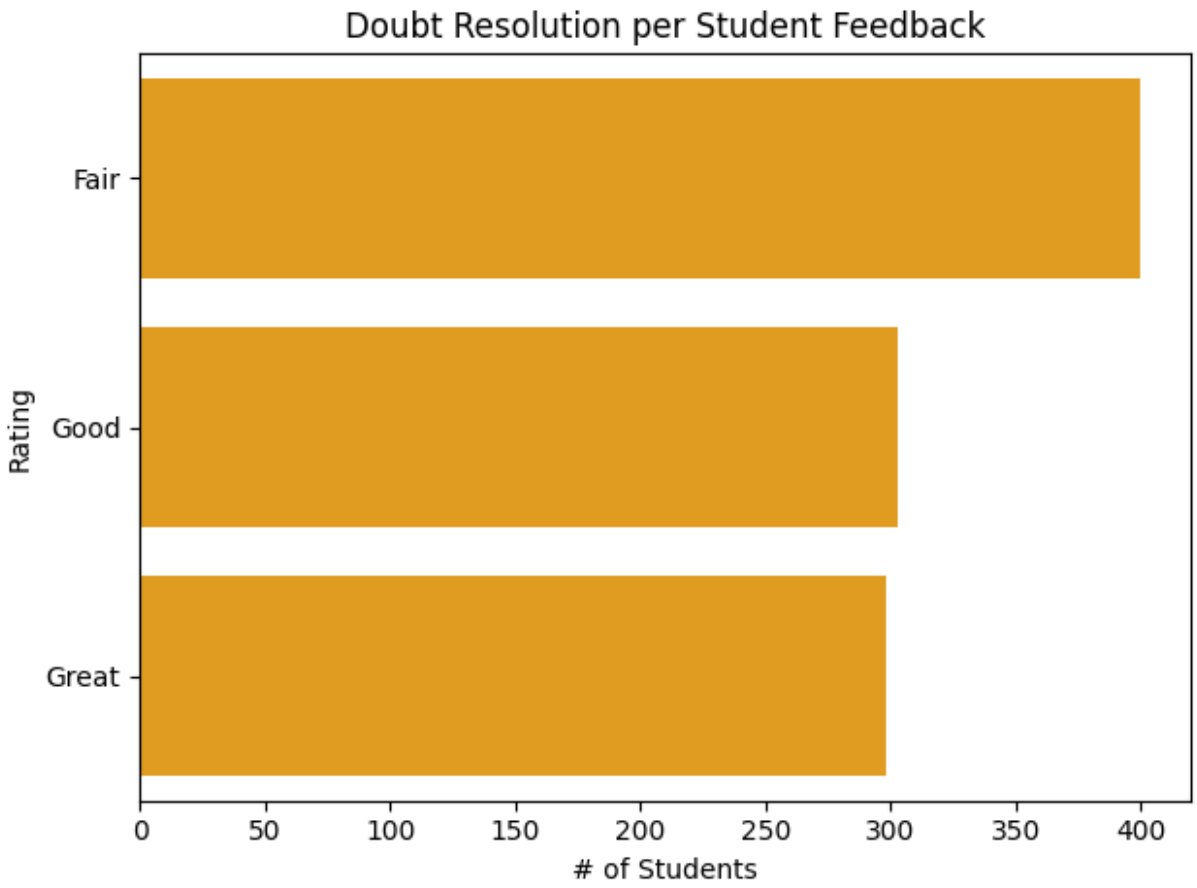
```
ax = sns.countplot(structuring_of_the_course_group, color='orange')
ax.set_title('Course Structure per Student Feedback')
ax.set_xlabel('# of Students')
ax.set_ylabel('Rating')
plt.tight_layout()
plt.show()

#Student Support rating per student feedback
provides_support_for_students_going_above_and_beyond_group =
pd.qcut(df.provides_support_for_students_going_above_and_beyond,labels
=bin_labels,q=3)
ax =
sns.countplot(provides_support_for_students_going_above_and_beyond_gro
up, color='orange')
ax.set_title('Student Support per Student Feedback')
ax.set_xlabel('# of Students')
ax.set_ylabel('Rating')
plt.tight_layout()
plt.show()

#Course recommendation rating per student feedback
course_recommendation_based_on_relevance_group =
pd.qcut(df.course_recommendation_based_on_relevance,labels=bin_labels,
q=3)
ax = sns.countplot(course_recommendation_based_on_relevance_group,
color='orange')
ax.set_title('Course Recommendation per Student Feedback')
ax.set_xlabel('# of Students')
ax.set_ylabel('Rating')
plt.tight_layout()
plt.show()
```

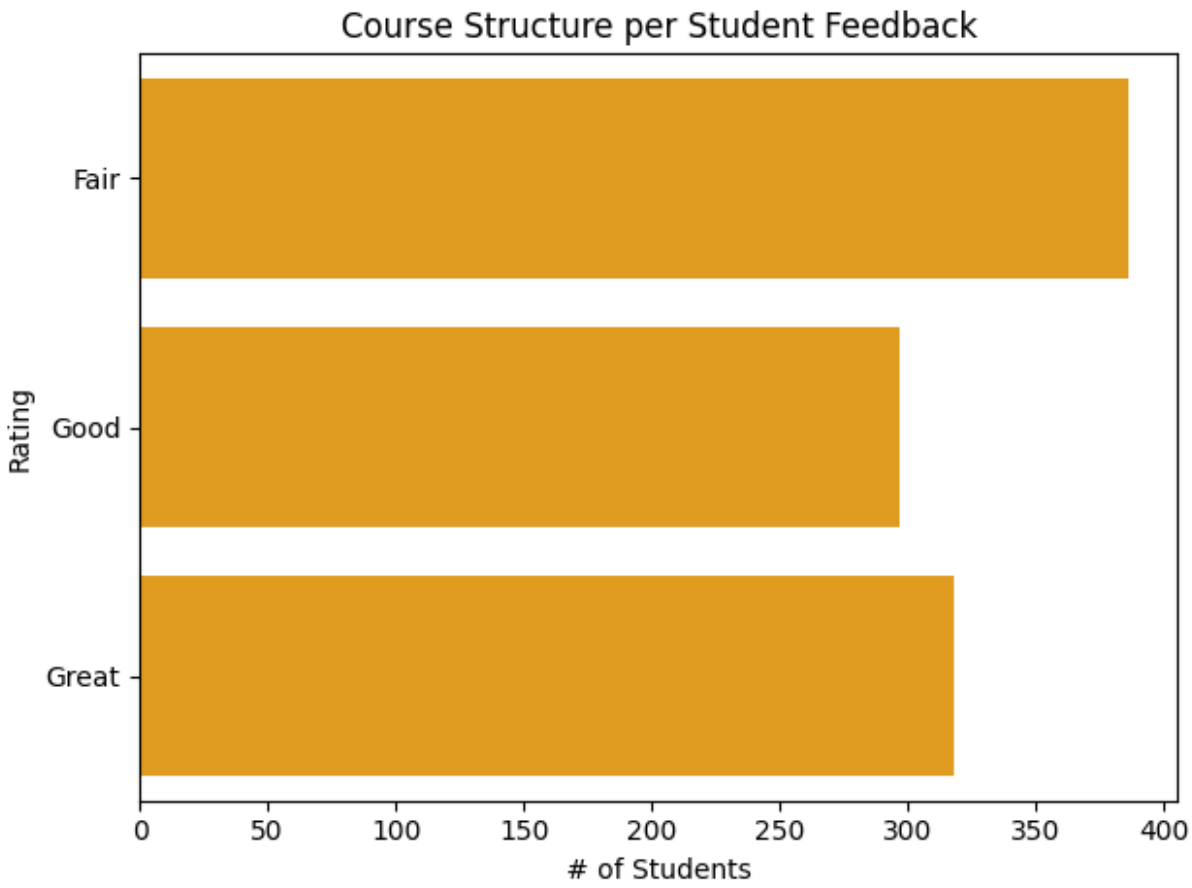

Knowledge Level per Student Feedback
Rating
Fair
Good
Great
0
100
200
300
400
500
# of Students

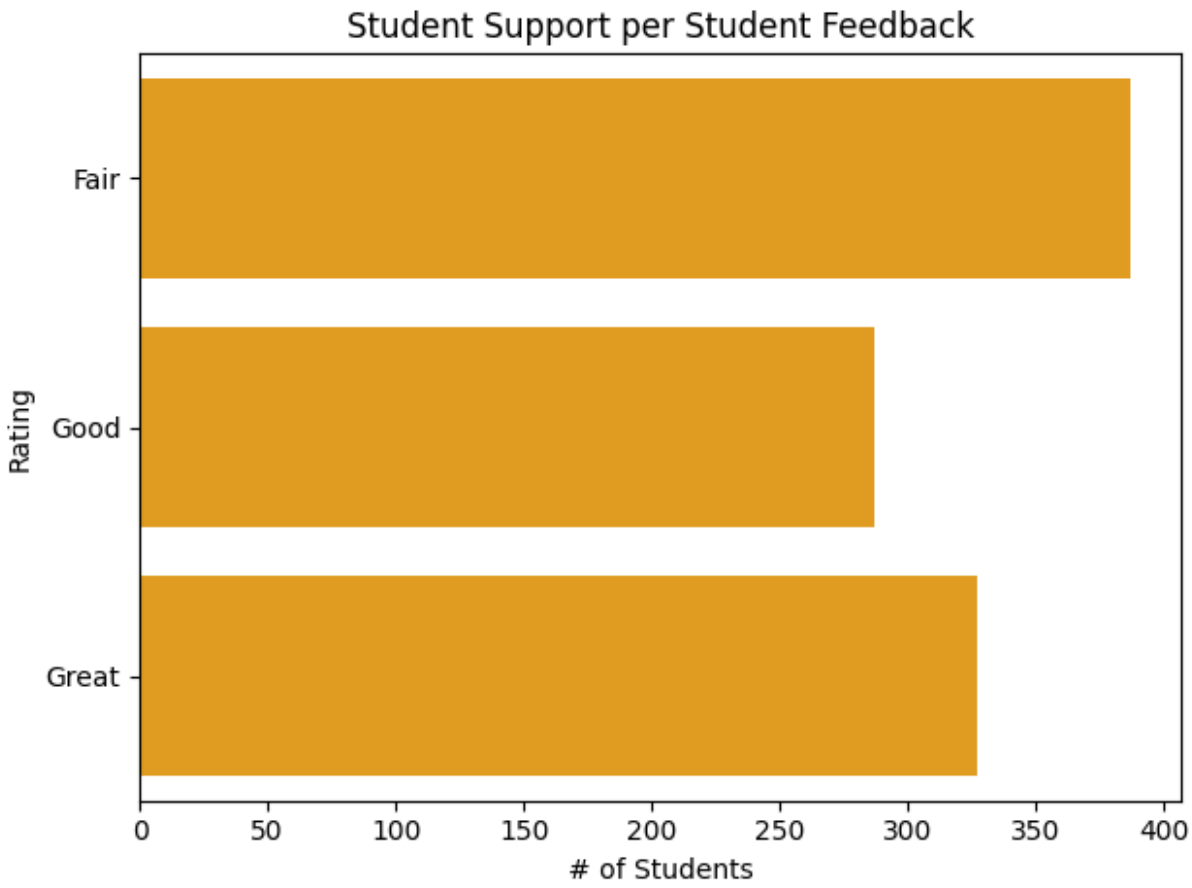

Presentation usage per Student Feedback
Rating
Fair
Good
Great
0
50
100
150
200
250
300
350
400
# of Students

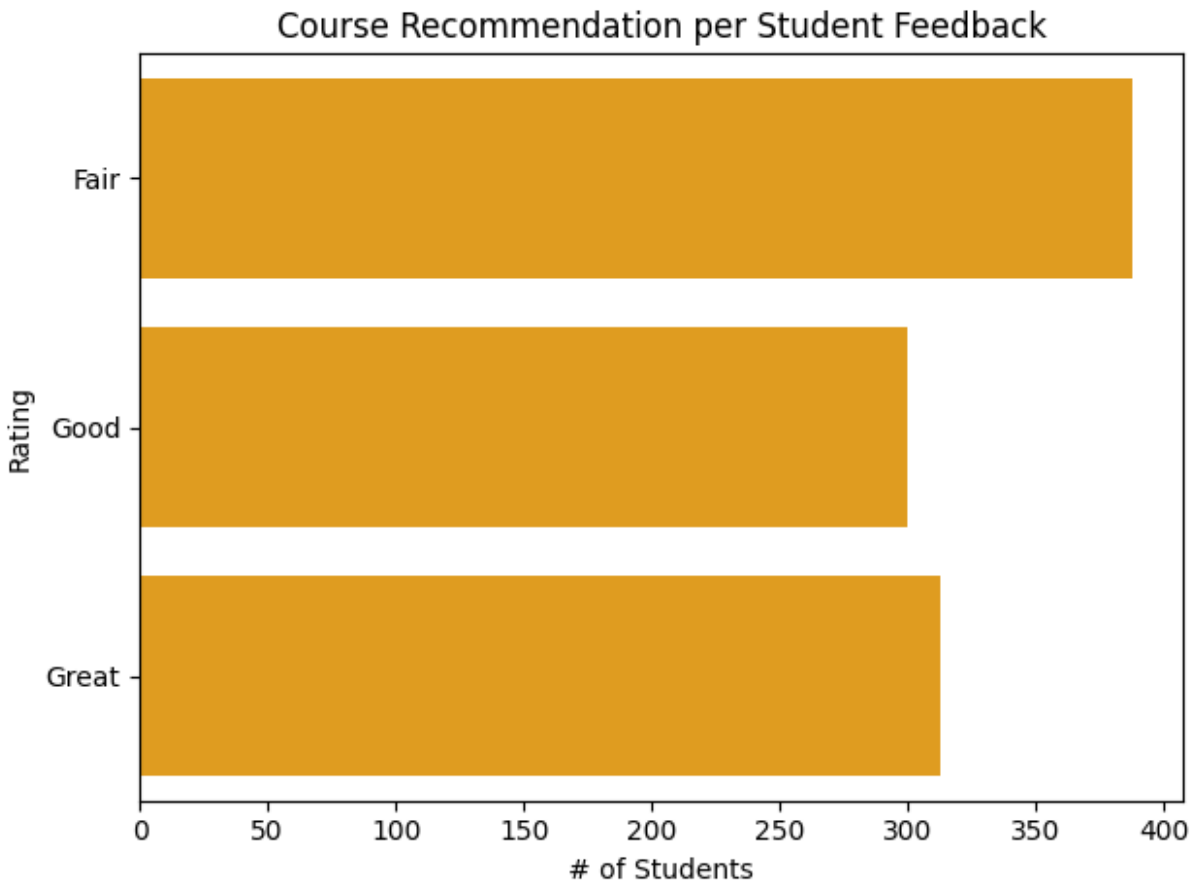

Concepts Explanation per Student Feedback
Rating
Fair
Good
Great
0
50
100
150
200
250
300
350
400
# of Students


Assignment Difficulty per Student Feedback
Rating
Fair
Good
Great
0
50
100
150
200
250
300
350
400
# of Students


Doubt Resolution per Student Feedback
Rating
Fair
Good
Great
0
50
100
150
200
250
300
350
400
# of Students


Course Structure per Student Feedback
Fair
Good
Great
Rating
0
50
100
150
200
250
300
350
400
# of Students


Student Support per Student Feedback
Rating
Fair
Good
Great
0
50
100
150
200
250
300
350
400
# of Students



5. Insights and Recommendations

From the analysis carried out above, we can uncover a few things:

- The teacher has a good knowledge of the subject matter but the course could be structured in a better manner.
- The teacher's presentation techniques, when explaining course concepts to students, could be better.
- The teacher could also take time to explain course concepts to students to facilitate their understanding.
- The teacher could incorporate harder assignments to students as a way to assess their learning curve and reinforce understanding.
- The teacher takes time to understand and clarify doubts students have about course concepts but there's always room for improvement.
- Student support by the teacher is generally fair so there is room for improvement.

Overall, the teacher is doing okay but there is plenty of room for improvement on teaching performance.

6. Export of clean data

The final step would be to export the cleaned data and features in .csv format

```
out_csv = "student_feedback_survey_cleaned.csv"
df.to_csv(out_csv, index=False)
print(f"Saved: {out_csv}")
```

```
Saved: student_feedback_survey_cleaned.csv
```